Total molecular photoionization cross-sections by algebraic diagrammatic construction-Stieltjes-Lanczos method: Benchmark calculations

M. Ruberti,1 R. Yun,1 K. Gokhberg,2 S. Kopelke,2 L. S. Cederbaum,2 F. Tarantelli,3 and V. Averbukh1, a)
1Department of Physics, Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom
2Theoretische Chemie, Physikalisch-Chemisches Institut, Universität Heidelberg, Im Neuenheimer Feld 229, D-69120 Heidelberg, Germany
3Dipartimento di Chimica, Università degli Studi di Perugia and CNR I.S.T.M., Via Elce di Sotto 8, 06123 Perugia, Italy

(Received 26 July 2013; accepted 23 September 2013; published online 10 October 2013)

In [K. Gokhberg, V. Vysotskiy, L. S. Cederbaum, L. Storchi, F. Tarantelli, and V. Averbukh, J. Chem. Phys. 130, 064104 (2009)] we introduced a new \$L^2\textit{ab initio}$ method for the calculation of total molecular photoionization cross-sections. The method is based on the \textit{ab initio} description of discretized photoionized molecular states within the many-electron Green’s function approach, known as algebraic diagrammatic construction (ADC), and on the application of Stieltjes-Chebyshev moment theory to Lanczos pseudospectra of the ADC electronic Hamiltonian. Here we establish the accuracy of the new technique by comparing the ADC-Lanczos-Stieltjes cross-sections in the valence ionization region to the experimental ones for a series of eight molecules of first row elements: HF, NH$_3$, H$_2$O, CO$_2$, H$_2$CO, CH$_4$, C$_2$H$_2$, and C$_2$H$_4$. We find that the use of the second-order ADC technique [ADC(2)] that includes double electronic excitations leads to a substantial systematic improvement over the first-order method [ADC(1)] and to a good agreement with experiment for photon energies below 80 eV. The use of extended second-order ADC theory [ADC(2)x] leads to a smaller further improvement. Above 80 eV photon energy all three methods lead to significant deviations from the experimental values which we attribute to the use of Gaussian single-electron bases. Our calculations show that the ADC(2)-Lanczos-Stieltjes technique is a reliable and efficient \textit{ab initio} tool for theoretical prediction of total molecular photo-ionization cross-sections in the valence region.

© 2013 AIP Publishing LLC. [http://dx.doi.org/10.1063/1.4824431]

I. INTRODUCTION

The molecular single-photon ionization cross-section is defined as the probability per unit time and per unit incident photon flux density of ionizing a molecule by absorbing a photon.\textsuperscript{1} Electron emission is a strongly dominant decay mechanism for molecules excited a few eV and higher above the ionization threshold, such that the photoabsorption and photoionization cross-sections become nearly coincident in this energy range. Here we consider photoionization within the vertical transition approximation, i.e., assuming that molecular geometry is fixed at the equilibrium geometry of the neutral and does not change in the course of the electronic transition. In this framework, the cross-section can be viewed as the probability of one-photon absorption leading to a state in the electronic continuum. In the case of randomly oriented molecules, the expression for the ground state photoabsorption cross-section in SI units is given by

\[
\sigma(E) = \frac{\pi e^2 \hbar}{2\varepsilon_0 m_c c} \frac{df(E)}{dE},
\]

where \(\frac{df(E)}{dE}\) is the oscillator strength density and \(E = \hbar \omega_0\) is the photon energy. Within the dipole approximation in the length gauge, the oscillator strength is given by the dipole matrix element between the ground state (\(\Psi_0\)) and the final continuum state (\(\Psi_E\)) of the N-electron system:

\[
\frac{df(E)}{dE} = \frac{2m_e E}{3\hbar^2} \left| \left\langle \frac{\Psi_E}{\Psi_0} \sum_{j=1}^{N} \mathbf{r}_j \right| \Psi_0 \right|^2,
\]

where the final continuum states are normalized to \(\delta\)-function in energy:

\[
\langle \Psi_E | \Psi_E \rangle = \delta(E - E').
\]

If more than a single state of the cation is accessible at the photon energy \(E\), an incoherent summation over all open ionization channels should be included into Eq. (2). The calculation of photoionization cross-sections requires the knowledge of many-electron wave functions belonging to the continuum part of the spectrum. The basic computational problem one faces here is taking into account both the scattering character of the photoionized state wave function and the electron correlation. While many well-developed theoretical techniques exist for the description of atomic photoionization,\textsuperscript{2} the multi-centre molecular problem still poses a formidable challenge to the theory. The state of the art theoretical methods for calculation of molecular photoionization cross-sections either do not take into account sufficiently the electronic correlation, see,

______________________________
\textsuperscript{a)Electronic mail: v.averbukh@imperial.ac.uk}
e.g., Ref. 3, or treat the photo-ionization continuum rather approximately, see, e.g., Ref. 4. Highly accurate many-electron wave functions and transition matrix elements are routinely obtained by the post-Hartree-Fock (post-HF) methods of \textit{ab initio} quantum chemistry.\textsuperscript{5} However, these methods are based on the use of finite sets of square-integrable (typically Gaussian) single-electron basis functions. As a result, the computed molecular eigenstates in the continuum energy region are discrete and \( L^2 \)-normalized. Such states cannot be used directly in the photoionization cross-section computation [see Eq. (1)]. Nevertheless, it turns out that the discretized continuum can be utilized for computing the spectral moments of the dipole oscillator strength (2). Mathematical technique of Stieltjes-Chebyshev moment theory or Stieltjes imaging (SI) can then be used to reconstruct the true photo-ionization cross-section from a finite series of moments.\textsuperscript{6}

The SI technique can be seen as a practical and mathematically well defined procedure for renormalization and interpolation of the oscillator strength density, starting from a general discretized spectrum formed by energies and oscillator strengths obtained from an \( L^2 \) calculation.

The main computational bottleneck of SI in its original formulation\textsuperscript{6} is the need to fully diagonalize the Hamiltonian of the system to extract the full spectrum of discretized final states. This effectively restricts the use of the technique to either small systems (e.g., atoms, diatomics) or to low-accuracy \textit{ab initio} approximations for the photoionized states (e.g., single-excitation schemes). Indeed, Hamiltonian matrix dimensions for polyatomic molecules represented using high-quality single-electron basis sets in computational schemes going beyond single excitations easily exceed the millions, making these Hamiltonians not amenable to full diagonalization. This drawback of the Stieltjes imaging technique was realized early on by Nesbet\textsuperscript{6} and a number of methods for overcoming this problem has been proposed since then.\textsuperscript{5–10} In Ref. 11 we described a new general approach for application of the SI technique to problems involving Hamiltonian matrices of large dimension. Our method is based on applying the SI procedure to a relatively small block-Lanczos pseudospectrum\textsuperscript{12–14} instead of the full spectrum of the molecular Hamiltonian. In our study we used a hierarchy of \textit{ab initio} methods of the algebraic diagrammatic construction (ADC) type,\textsuperscript{15} our proof-of-concept applications included photoionization cross-sections of He, Ne, and benzene\textsuperscript{11} (we have also shown that a similar technique can be used for the SI computations of the decay widths\textsuperscript{16}). Very recently, an analogous Lanczos-Stieltjes approach has been developed for the coupled cluster pseudospectra by Cukras \textit{et al}.\textsuperscript{17} Here we present a systematic study of the accuracy of the ADC-Lanczos-Stieltjes technique using a test set of eight molecules of first-row elements, for which high-quality experimental cross-sections are available in the literature.

This article is organized as follows. The relevant aspects of the ADC approach to electronic excited states is presented in Sec. II. The Lanczos iterative diagonalization method is briefly reviewed in Sec. III. Section IV is devoted to testing the accuracy of the molecular photoionization cross-sections obtained by SI of block-Lanczos pseudospectra of a hierarchy of ADC Hamiltonians. Conclusions are given in Sec. V.

II. ADC-STIELTJES-LANCZOS METHOD FOR MOLECULAR PHOTOIONIZATION CROSS-SECTIONS

A. ADC \textit{ab initio} schemes within the intermediate state representation

The ADC schemes for excited states of closed-shell systems were originally derived as approximations to the polarization propagator, based on an algebraic reformulation of its diagrammatic perturbation theory. The ADC(n) polarization propagator is complete up to order \( n \) of perturbation theory in the electron interaction and includes also higher-order diagrams in the form of infinite partial (incomplete) summations.\textsuperscript{18} It was later recognized\textsuperscript{19} as being interpretable as a wave-function method as well. In fact, ADC establishes a connection between propagator and wave-function methods. The latter interpretation comes from the explicit construction of the intermediate states representation (ISR) that gives rise to the ADC form of the propagator, providing an alternative approach to the hierarchy of the ADC schemes.\textsuperscript{19, 20} The starting point is the construction of the so-called correlated excited states (CES), defined as

\[ |\Psi_i^0\rangle = \hat{\mathcal{C}}_i^\dagger |\Psi_0\rangle, \]

where the operators \( \hat{\mathcal{C}}_i^\dagger \) denote the physical excitation operators corresponding respectively to 1p1h, 2p2h etc. excitations,

\[ \hat{\mathcal{C}}_i^\dagger = \{ \hat{a}_{a_i}^\dagger \hat{a}_{b_i} \hat{a}_{a_j}^\dagger \hat{a}_{b_j}^\dagger \hat{a}_{a_k} \hat{a}_{b_k} \ (a < b, j < k) \ldots \}, \]

and \( |\Psi_0\rangle \) is the exact correlated ground state of the system.

This non-orthogonal CES basis set is complete in the space of the excited states of the N-electron system\textsuperscript{21} and has the advantage that ground state correlation is already built into every basis vector. It can be orthonormalized in a two-step procedure. First one performs Gram-Schmidt orthogonalization of each excitation class with respect to all the lower excitation classes. The states \( |\Psi_{m}^{p}\rangle \) formed in this first step are referred to as precursor states. The second step is symmetric orthonormalization of the resulting precursor states within each excitation class. As an example the procedure for the first (1h1p) excitation class gives the following precursor states:

\[ |\Psi_{ai}^{1}\rangle = \hat{\mathcal{A}}_{a_i}^\dagger |\Psi_0\rangle - |\Psi_0\rangle \langle \hat{\mathcal{A}}_{b_i}^\dagger \hat{\mathcal{A}}_{a_i}^\dagger |\Psi_0\rangle. \]

The second step gives

\[ |\bar{\Psi}_{ai}^{1}\rangle = \sum_{bj}|\Psi_{bj}^{1}\rangle |S_{bj,ai}^{-\frac{1}{2}}\rangle, \]

where \( S \) is the overlap matrix of the first excitation class precursor states, i.e.,

\[ S_{bj,ai} = \langle \Psi_{bj}^{1}|\Psi_{ai}^{1}\rangle. \]

In a compact notation the excitation class orthogonalized (ECO) states can be written as

\[ |\tilde{\Psi}_x^m\rangle = \hat{\mathcal{O}}_{x}^{m-1} \sum_y |\Psi_y^m\rangle \left(S_{yx}^m\right)^{-\frac{1}{2}}, \]
where $S_{xy}^{m}$ is defined as
\[ S_{xy}^{m} = \langle \hat{Q}_{y}^{m} \mid \hat{Q}_{x}^{m-1} \rangle \] (10)
and
\[ \hat{Q}_{x}^{m} = \hat{1} - \sum_{l=0}^{m} \hat{P}_{l} \] (11)
is the projector operator onto the space orthogonal to the first $m$ excitation classes. Finally, every intermediate state can be expressed as
\[ \mid \tilde{\Psi}_{I} \rangle = \tilde{C}_{I} \mid \Psi_{0} \rangle, \] (12)
where all the effects of the consecutive orthonormalizations are encoded in the new creation operators $\tilde{C}_{I}$.

The ADC secular matrix is the representation of the shifted electronic Hamiltonian operator $\hat{H} - E_{0}$ in the ECO-CES space:
\[ \mathcal{H}_{I} = \langle \tilde{\Psi}_{I} \mid \hat{H} - E_{0} \mid \tilde{\Psi}_{J} \rangle = \langle \tilde{\Psi}_{0} \mid \tilde{C}_{I} \left[ \hat{H}, \tilde{C}_{J} \right] \mid \tilde{\Psi}_{0} \rangle. \] (13)
At this point Møller-Plesset (MP) perturbation theory is introduced to describe the ground state correlation, i.e., $\mid \Psi_{0} \rangle$ and $E_{0}$:
\[ \mid \Psi'_{0} \rangle = \mid \Phi_{0}^{HF} \rangle + \mid \Psi_{0}^{[1]} \rangle + \mid \Psi_{0}^{[2]} \rangle + \mid \Psi_{0}^{[3]} \rangle + \ldots, \] (14)
in which the first order correction $\mid \Psi_{0}^{[1]} \rangle$ contains only double excitations (2h2p) with respect to $\mid \Phi_{0}^{HF} \rangle$, while $\mid \Psi_{0}^{[2]} \rangle$ contains single, double, triple, and quadruple excitations.

The vertical excitation energies are obtained solving the eigenvalue problem $HV = \omega V$ and the excited eigenstates of the system are, therefore, given on the basis of the intermediate states:
\[ \mid \Psi_{n} \rangle = \sum_{I} V_{I,n} \mid \tilde{\Psi}_{I} \rangle. \] (15)
Having this explicit expression for the excited states of the system the transition moments of the type $\langle \Psi_{m} \mid \hat{D} \mid \Psi_{0} \rangle$ are given by
\[ \langle \Psi_{m} \mid \hat{D} \mid \Psi_{0} \rangle = V_{m}^{t} \cdot F = \sum_{rs} d_{rs} V_{m}^{t} \cdot f_{rs}, \] (16)
where $d_{rs}$ are the matrix elements of the dipole operator on the basis of the one particle orbitals. The matrix of effective transition amplitudes $F$ and the associated vector $V_{m}$ are defined correspondingly by
\[ f_{1,rs} = \langle \tilde{\Psi}_{1} \mid \hat{a}_{s}^{\dagger} \hat{a}_{r} \mid \tilde{\Psi}_{0} \rangle, \quad F_{l} = \langle \tilde{\Psi}_{l} \mid \hat{D} \mid \tilde{\Psi}_{0} \rangle. \] (17)

The hierarchy of ADC(n) approximations is obtained for each order $n$ by truncating the intermediate state manifold at some limiting excitation class and, also, by truncating the resulting perturbation expansions for the included classes in a way consistent with the polarization propagator approach. For example, at the ADC(2) level the matrix in the whole space of the system can be represented as
\[ \mathcal{H}_{ADC}^{(2)} = \mathcal{H}_{0}^{(0)} + \mathcal{H}_{1.1}^{(1.1)} + \mathcal{H}_{1.1}^{(1.2)} + \mathcal{H}_{1.2}^{(1.1)} + \mathcal{H}_{2.1}^{(1.1)}. \] (18)

Thus, in ADC(2) the perturbation expansion of the secular matrix elements extends through second, first, and zeroth order in the 1h1p block, the 1h1p-2h2p coupling block and the diagonal 2h2p block, respectively. In a similar way the 1h1p and 1h2p parts of the effective transition amplitudes have perturbation expansions through second and first order, respectively. An extension of the ADC scheme, not strictly consistent with the polarization propagator and referred to as ADC(2) extended [ADC(2)x], is obtained by using the first order expansion for the 2h2p block that accounts for the couplings between the 2h2p intermediate states. The ADC(n) schemes are size consistent and compact relative to the corresponding truncated CI expansions.

B. Stieltjes-Chebyshev moment theory

The SI approach to photoionization cross-sections is based on the assumption that the discretized $\mathcal{L}^{2}$ spectrum above the ionization threshold allows one to obtain good approximations for lower spectral moments, $S(n)$:
\[ S(n) = \langle \Psi_{0} \mid \hat{D} \hat{N}^{n} \hat{D} \mid \Psi_{0} \rangle. \] (19)
It is possible to express the moments (19) in terms of the exact bound $\{ \mid \Psi_{j} \rangle \}$ and continuum $\{ \mid \Psi_{E} \rangle \}$ eigenstates of the Hamiltonian using the resolution of identity:
\[ S(n) = \sum_{j} E_{j}^{n+1} \langle \Psi_{0} \mid \hat{D} \mid \Psi_{j} \rangle^{2} \]
\[ + \int_{E_{\text{threshold}}}^{\infty} E^{n+1} \langle \Psi_{0} \mid \hat{D} \mid \Psi_{E} \rangle^{2} dE. \] (20)
Assuming that the function $\hat{D} \mid \Psi_{0} \rangle$ is non-zero only within a finite interaction region, one can express the spectral moments using the resolution of identity in terms of the variationally calculated discrete $\mathcal{L}^{2}$ eigenstates $\{ \mid \Phi_{a} \rangle \}$ spanning the same finite region:
\[ S(n) \approx \sum_{a} E_{a}^{n+1} \langle \Psi_{0} \mid \hat{D} \mid \Phi_{a} \rangle^{2}. \] (21)
The moment problem consists in recovering the continuous function $E \langle \Psi_{0} \mid \hat{D} \mid \Psi_{E} \rangle^{2}$ from a finite number of its spectral moments obtained within the $\mathcal{L}^{2}$ approximation (21). It is important to note that within the non-relativistic theory the spectral moments diverge for $n > 1$ and thus the SI approach must rely on the use of negative spectral moments.

The SI computational procedure includes construction of the quadrature pseudospectrum defined by the following 2n equations:
\[ S(-2k) = \sum_{i=1}^{n} f_{i}^{(n)} \left[ \xi_{i}^{(n)} \right]^{2k}, \quad k = 1, 2, \ldots, 2n. \] (22)
Thus, the first 2n moments reconstructed from the $\mathcal{L}^{2}$ calculation give rise to an n-term smoothed ($n \ll N$) principal pseudospectra $\{ \xi_{i}^{(n)}, f_{i}^{(n)} \}$ 1 = 1, ..., n. These quadrature pseudospectra are then used to determine the cumulative oscillator strength, $F^{(n)}(\epsilon)$, which is a histogram-like multi-step function. The cumulative oscillator strength obtained in this way approximates the exact one, converging to it for large values of $n$ if the first 2n moments are accurately reproduced by the $\mathcal{L}^{2}$ calculation. This is formally expressed by the Chebyshev relation which shows that $F^{(n)}(\epsilon)$ gives upper
and lower bounds on $F(\varepsilon)$ at the points $\varepsilon_i^{(n)}$. The SI approximation to the oscillator strength is found by differentiating the Stieltjes-Chebyshev cumulative function, according to the Stieltjes derivative definition (different from the Dirac convention for stepwise functions).

III. CALCULATION OF SPECTRAL MOMENTS USING LANCZOS DIAGONALIZATION PROCEDURE

Direct application of the SI procedure requires full diagonalization of the many-electron [e.g., ADC(n)] Hamiltonian matrix. This is typically not feasible for \textit{ab initio} methods going beyond single electronic excitations, e.g., ADC(2). In Ref. 11, we have proposed to overcome this difficulty by applying the SI procedure to the relatively low-dimension block-Lanczos pseudospectrum of the full ADC(n) Hamiltonian. Within the Lanczos method,12–14 the Hamiltonian is represented on the basis of the so-called Lanczos states, $|\psi_j\rangle$, which are obtained by Gram-Schmidt orthogonalization of the Krylov states, $|\varphi_k\rangle = \hat{H}^k|\varphi_0\rangle$, $k = 0, 1, 2, \ldots, N - 1$. The initial state $|\varphi_0\rangle$ is usually chosen to have maximal overlap with the Hamiltonian eigenstates that are of interest in the given physical problem. The Lanczos states of successive orders can be used to construct a series of approximations to the Hamiltonian. The $N$th order Lanczos approximation to $\hat{H}$ is of the form

$$\hat{H}^{(N)} = \sum_{j,k=0}^N |\Phi_j\rangle\langle \Phi_j|\hat{H}|\Phi_k\rangle\langle \Phi_k|,$$

and is a tridiagonal matrix, i.e., $\langle \Phi_j|\hat{H}|\Phi_k\rangle = 0$ for $|j - k| > 1$. The eigenvalues and eigenvectors of the operator $\hat{H}^{(N)}$ form Lanczos pseudospectrum. With increasing $N$, the Lanczos pseudospectrum becomes a successively better approximation to the spectrum of $\hat{H}$. A generalization of the Lanczos technique to the case of a set of initial states is called block-Lanczos method.13 The Lanczos (block-Lanczos) method is useful not only for diagonalizing Hamiltonian matrices of large dimensions, but can also be used for the calculation of the spectral moments of the type of (21). To this end, the original Hamiltonian has to be substituted by its Lanczos representation:

$$S_n \approx S_n^{(N)} = \sum_{a=0}^N (E_a^{(N)})^n \left| \langle \Psi_0|\hat{D}|\chi_a^{(N)} \rangle \right|^2.$$  

While the non-negative moments of the order $0 \leq n \leq 2N$ can be calculated exactly by Eq. (24) with the appropriate choice of the initial state (or block of states),13 the negative moments required for the Stieltjes imaging can be calculated only approximately. Since the Lanczos (block-Lanczos) algorithm approximates most effectively the eigenstate subspace spanned by the starting vector (or vectors), one has to consider the physical properties of the system when choosing the initial guess. In the case of photoionization cross-section the final states of the process that we wish to resolve are of 1h1p type and have the symmetry of the $D|\Psi_0\rangle$ state. Therefore, we choose our starting vectors for the block-Lanczos iterations as the full set 1h1p ADC intermediate states of the appropriate symmetry.

IV. MOLECULAR PHOTOIONIZATION CROSS-SECTIONS BY ADC-LANCZOS-STIETJES METHOD

The primary goal of this work is to test the accuracy of the ADC-Lanczos-Stieltjes method at the ADC(1), ADC(2), and ADC(2)x levels of \textit{ab initio} theory using a test set of molecules for which accurate experimental total ionization cross-sections are available. Construction of the ADC Hamiltonian matrices requires carrying out restricted HF calculations and transforming the electron repulsion integrals from the atomic orbital (AO) basis to the molecular orbital (MO) basis. Throughout this work, these tasks are performed using MOLCAS 7.6 quantum chemical program package.24 The standard Gaussian basis sets used in the present study have been obtained from Ref. 25. Kaufmann-Baumeister-Jungen (KBJ) continuum-like diffuse Gaussian functions26 are systematically used in each calculation to augment the standard basis sets for a strictly necessary better representation of the discretized electronic continuum.

We restrict the calculated cross-sections to the energy range of up to 100 eV, thus including valence-type excitations but excluding excitations of the core electrons. Stieltjes imaging procedure is carried out in quadruple precision using the algorithm of Ref. 7. The presented photoionization cross-sections are obtained as interpolation of the discrete points corresponding to several (up to a maximum of five) successive Stieltjes orders for which approximate stationarity of the results is achieved. The interpolation procedure analyzes the points generated by each Stieltjes order and finds which orders satisfy the stationarity condition within a predetermined maximal deviation. The procedure then merges the cross-section values belonging to different orders in the stationarity region into a single data set and finally performs the merged data set interpolation. For all of the molecules studied in the present work, the approximate stationarity is reached for several consecutive Stieltjes orders between $n = 5$ and $n = 15$. This gives an idea of the number of principal representation states and the energy resolution it is possible to achieve with the method presented.

We quantify the deviation of the ADC-Lanczos-Stieltjes cross-sections from the experimental ones computing their energy-dependent and energy-averaged relative discrepancies over the covered photon energy region. For all systems we have used correlation consistent basis sets of the cc-pCVnZ type, with $n = T$ or $n = Q$ depending on the system and on the specific atom within the molecule; we have seen that the inclusion of large exponents for the accurate description of the localized core electrons can lead to a significant improvement in the cross-section. In some cases we have noticed that uncontracting the basis set leads to an improvement in the cross-section, while in other cases it did not make substantial difference. For every molecule we have checked the convergence of the results of the spectral moments and of the cross-section, with respect to the choice of the details of the basis sets, i.e., with respect to the choice of the number of KBJ
The converged cross-sections have been obtained for the dimensions of the Lanczos pseudospectra for which the convergence has been achieved. In Table I we report the details of the ADC-Lanczos-Stieltjes photoionization cross-section calculations: the basis sets, the dimensions of the ADC(2) matrices, and the dimensions of the Lanczos pseudospectra for which the converged cross-sections have been obtained.

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Basis set</th>
<th>ADC(2) matrix dim.</th>
<th>Krylov space dim.</th>
</tr>
</thead>
<tbody>
<tr>
<td>H$_2$O</td>
<td>O:cc-pCVQZ + (10s10p4d); H:cc-pVQZ + (5s5p5d)</td>
<td>157 653</td>
<td>3390</td>
</tr>
<tr>
<td>HF</td>
<td>F:UN-cc-pCVQZ + (10s10p10d); H:UN-cc-pVQZ + (6s6p6d2f)</td>
<td>181 236</td>
<td>3800</td>
</tr>
<tr>
<td>NH$_3$</td>
<td>N:UN-cc-pCVTZ + (5p5d); H:UN-cc-pVTZ + (3s3p)</td>
<td>121 376</td>
<td>3920</td>
</tr>
<tr>
<td>CH$_4$</td>
<td>C:UN-cc-pCVQZ + (6s6p6d2f); H:UN-cc-pVTZ + (4s4p2d)</td>
<td>209 682</td>
<td>3980</td>
</tr>
<tr>
<td>C$_2$H$_2$</td>
<td>C:cc-pCVTZ + (10s10p10d); H:cc-pVTZ + (6s6p6d2f)</td>
<td>252 025</td>
<td>3170</td>
</tr>
<tr>
<td>C$_2$H$_4$</td>
<td>C:cc-pCVTZ + (7s10p10d); H:cc-pVTZ + (3s3p3d)</td>
<td>411 931</td>
<td>3870</td>
</tr>
<tr>
<td>CH$_3$OH</td>
<td>C:O:cc-pCVTZ + (2s4p3d); H:UN-cc-pVTZ + (1s1p)</td>
<td>182 194</td>
<td>3960</td>
</tr>
<tr>
<td>CO$_2$</td>
<td>C:O:cc-pCVTZ + (5s6p6d2f)</td>
<td>391 838</td>
<td>3920</td>
</tr>
</tbody>
</table>

Exponents and the quality of the cc-type basis set; we report in the present work only the results obtained using the basis set at which the convergence has been achieved. In Table I we denote a fully uncontracted basis set by the prefix UN.

Experimental total photo-ionization cross-section as well as a series of SI results obtained via full diagonalization of the ADC(1) matrix and block Lanczos diagonalization of the ADC(2) and ADC(2)x matrices are reported in Figs. 1–8 for the molecules considered. The basis set employed, as well as the dimension of the ADC(2)x matrices and that of the Lanczos pseudospectrum space for which the convergence in the cross-section has been obtained are reported in Table I. The average relative deviations are reported in Table II.

### A. Water

Figure 1 shows the experimental total photoionization cross section of the H$_2$O molecule as well as a series of Stieltjes imaging results obtained via full diagonalization of the ADC(1) matrix and block-Lanczos diagonalization of the ADC(2) and ADC(2)x Hamiltonian matrices. The details of the calculations can be found in Table I. One can see that the agreement between the experimental and the theoretical cross sections improves dramatically from ADC(1) to ADC(2) level, but not nearly as much when going from ADC(2) to the ADC(2)x level. The experimental measurement in Ref. 27 were performed using the dipole (e,2e) electron scattering technique, in the more recent study 29 the dipole (e,e) spectroscopy technique has been used, while in Ref. 28 the cross-section is measured directly with photoabsorption techniques, by use of a double ionization chamber. The ADC(2) and the ADC(2)x results essentially coincide with the newer experimental data 28,29 (apart from the sharp feature at 15 eV), but show visible deviation from the older experimental results around 20–30 eV. 27 In Table II we cite relative deviations of the theoretical results from the newer experimental data.
B. Hydrogen fluoride

Figure 2 shows the experimental total photoionization cross section of the HF molecule as well as a series of Stieltjes imaging results obtained via full diagonalization of the ADC(1) matrix and block-Lanczos diagonalization of the ADC(2) and ADC(2)x Hamiltonian matrices. The details of the calculations can be found in Table I. The most recent experimental result we have found is from 1981 and the experimental method used in that work is the magic-angle dipole (e,2e) spectroscopy technique. Our results show that, although the higher-order schemes lead to a much improved peak.

FIG. 3. Total photoionization cross-section of NH₃. Squares – experimental result of Ref. 32, crosses – experimental result of Ref. 33, dashed-dotted line – ADC(1)-Stieltjes result, dashed line – ADC(2)-Lanczos-Stieltjes cross-section obtained using block-Lanczos pseudospectrum of 3920 eigenvalues and eigenvectors, full line – ADC(2)x-Lanczos-Stieltjes cross-section obtained using block-Lanczos pseudospectrum of 3920 eigenvalues and eigenvectors.

FIG. 4. Total photoionization cross-section of CH₄. Squares – experimental result of Ref. 34, dashed-dotted line – ADC(1)-Stieltjes result, dashed line – ADC(2)-Lanczos-Stieltjes cross-section obtained using block-Lanczos pseudospectrum of 3980 eigenvalues and eigenvectors, full line – ADC(2)x-Lanczos-Stieltjes cross-section obtained using block-Lanczos pseudospectrum of 3980 eigenvalues and eigenvectors.

FIG. 5. Total photoionization cross-section of C₂H₂. Squares – experimental results of Ref. 36, dashed-dotted line – ADC(1)-Stieltjes result, dashed line – ADC(2)-Lanczos-Stieltjes cross-section obtained using block-Lanczos pseudospectrum of 3170 eigenvalues and eigenvectors, full line – ADC(2)x-Lanczos-Stieltjes cross-section obtained using block-Lanczos pseudospectrum of 3170 eigenvalues and eigenvectors.

FIG. 6. Total photoionization cross section of C₂H₄. Squares – experimental results of Ref. 37, dashed-dotted line – ADC(1)-Stieltjes result, dashed line – ADC(2)-Lanczos-Stieltjes cross-section obtained using block-Lanczos pseudospectrum of 3870 eigenvalues and eigenvectors; full line – ADC(2)x-Lanczos-Stieltjes cross-section obtained using block-Lanczos pseudospectrum of 3870 eigenvalues and eigenvectors.
position, they do not reproduce precisely the cross-section peak height and the cross-section shoulder above 40 eV (see Table II for the average relative deviations of the computed hydrogen fluoride cross-sections from the experimental one).

We observe, however, that our theoretical results are in good agreement with the previous theoretical multichannel random phase approximation (MC-RPA) results of Cacelli et al. and the most recent coupled cluster results of Cukras et al. This may call for a revision of the experimental cross-section for HF.

C. Ammonia

Figure 3 shows the experimental total photoionization cross section of the NH$_3$ molecule as well as a series of Stieltjes imaging results obtained via full diagonalization of the ADC(1) matrix and block-Lanczos diagonalization of the ADC(2) and ADC(2)x Hamiltonian matrices. The details of the calculations can be found in Table I. As in the case of water and hydrogen fluoride, one can see that the agreement between the experimental and the theoretical cross sections improves with the order of the ADC scheme, with the difference between ADC(1) and ADC(2) being more critical than between ADC(2) and ADC(2)x. The experimental measurements in Ref. were performed using the dipole (e,e) spectroscopy technique, while in Ref. 32 the cross-section is measured directly by use of a double ionization chamber. Average relative deviations of the computed ammonia cross-sections from the experimental one are given in Table II. The main contribution to the deviation comes from the tail of the cross-section in the energy range above 60 eV.

D. Methane

Figure 4 shows the experimental total photoionization cross section of the CH$_4$ molecule as well as a series of Stieltjes imaging results obtained via full diagonalization of the ADC(1) matrix and block-Lanczos diagonalization of the ADC(2) and ADC(2)x Hamiltonian matrices. The details of the calculations can be found in Table I. The experimental measurement in Ref. 34 were performed using the dipole (e,e) spectroscopy technique. One can see that the agreement between the experimental and the theoretical cross sections improves with the order of the ADC scheme. The highest-order ADC(2)x result essentially coincides with the experimental one apart from the value at the peak that is underestimated by about 5 Mb. This underestimation of the maximum height of the peak is characteristic to some degree of the computed photo-absorption cross-sections of acetylene and ethylene (see below). Average relative deviations of the computed

<table>
<thead>
<tr>
<th></th>
<th>C$_2$H$_4$ (%)</th>
<th>C$_2$H$_2$ (%)</th>
<th>CH$_4$ (%)</th>
<th>CO$_2$ (%)</th>
<th>CH$_2$O (%)</th>
<th>H$_2$O (%)</th>
<th>HF (%)</th>
<th>NH$_3$ (%)</th>
<th>Average (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADC(1)</td>
<td>19.3</td>
<td>29.5</td>
<td>27.5</td>
<td>17.8</td>
<td>24.9</td>
<td>16.8</td>
<td>8.1</td>
<td>22.7</td>
<td>20.5</td>
</tr>
<tr>
<td>ADC(2)</td>
<td>15.8</td>
<td>18.6</td>
<td>22.5</td>
<td>7.0</td>
<td>17.5</td>
<td>7.9</td>
<td>7.2</td>
<td>19.0</td>
<td>14.0</td>
</tr>
<tr>
<td>ADC(2)x</td>
<td>12.4</td>
<td>14.6</td>
<td>17.7</td>
<td>7.1</td>
<td>16.4</td>
<td>7.8</td>
<td>7.6</td>
<td>17.6</td>
<td>12.1</td>
</tr>
</tbody>
</table>
methane cross-sections from the experimental one are given in Table II.

E. Acetylene

Figure 5 shows the experimental total photoionization cross section of the C2H2 molecule as well as a series of Stieltjes imaging results obtained via full diagonalization of the ADC(1) matrix and block-Lanczos diagonalization of the ADC(2) and ADC(2)x Hamiltonian matrices. The details of the calculations can be found in Table I. Acetylene photoionization has been well studied theoretically, in particular sharp resonance features in the fixed-geometry valence cross-sections has been revealed.35 The limited resolution of the SI procedure does not allow to reproduce such fine structures, however, due to vibrational broadening the experimental values can be directly compared to the SI results. Among the available experimental data, the most recent and extended set has been reported by Cooper et al.,36 who deduced it using dipole (e,e) and (e,e-ion) spectroscopies and we have chosen these data to compare our theoretical cross-section with. The same experimental method has been used by Cooper et al. to measure the photo-absorption cross-sections of ethylene and methane. One can see that the agreement between the experimental and the theoretical cross sections improves with the order of the ADC scheme. In particular, the position of the cross-section main peak at 15.5 eV is reproduced essentially exactly by the ADC(2)x scheme. The height of the peak is smaller with respect to the experimental measured value, being underestimated by about 5 Mb as in the case for methane. This difference is due to the energy resolution of the SI procedure. For the same reason also the double hump structure in the energy region between 13 eV and 16 eV is only approximately reproduced by our SI cross-section, resulting in a small shoulder at 13 eV. Average relative deviations of the computed acetylene cross-sections from the experimental one are given in Table II.

F. Ethylene

Figure 6 shows the experimental total photoionization cross section of the ethylene molecule as well as a series of Stieltjes imaging results obtained via full diagonalization of the ADC(1) matrix and block-Lanczos diagonalization of the ADC(2) and ADC(2)x Hamiltonian matrices. The details of the calculations can be found in Table I. The experimental measurement in Ref. 37 were performed using the dipole (e,e) spectroscopy technique. One can see that the agreement between the experimental and the theoretical cross sections improves significantly from the ADC(1) to the ADC(2) scheme. The highest-order ADC(2)x result, while more accurate in the tail region from 30 eV to 100 eV, does not improve the agreement with the experimental one in the main peak region. As in the case of acetylene, also the ethylene cross-section exhibits a sharp double hump structure in the energy region around 18 eV at the top of the peak, which is missed by both ADC(2) and ADC(2)x and is attributed to the resolution of the SI. On the contrary, the maximum value of the measured cross-section (≈61 Mb) is better reproduced, comparing to methane and acetylene. Average relative deviations of the computed ethylene cross-sections from the experimental one are given in Table II.

G. Formaldehyde

Figure 7 shows the experimental total photoionization cross section of the H2CO molecule as well as a series of Stieltjes imaging results obtained via full diagonalization of the ADC(1) matrix and block-Lanczos diagonalization of the ADC(2) and ADC(2)x Hamiltonian matrices. The details of the calculations can be found in Table I. The experimental measurements in Ref. 38 were performed using the dipole (e,e) spectroscopy technique. One can see that the agreement between the experimental and the theoretical cross sections improves with the order of the ADC scheme. Both ADC(2) and ADC(2)x methods struggle to reproduce the very sharp peak near 13 eV and the structure of the main peak at 18–20 eV, but give a better representation of the tail of the cross-section than the ADC(1) result. Average relative deviations of the computed formaldehyde cross-sections from the experimental one are given in Table II. As in the case of ammonia, the main contribution to the average deviation comes from the tail of the cross-section, in the energy range above 30 eV.

H. Carbon dioxide

Figure 8 shows the experimental total photoionization cross section of the CO2 molecule as well as a series of Stieltjes imaging results obtained via full diagonalization of the ADC(1) matrix and block-Lanczos diagonalization of the ADC(2) and ADC(2)x Hamiltonian matrices. The details of the calculations can be found in Table I. The experimental measurements in Ref. 39 were performed using the dipole (e,e) spectroscopy technique. One can see that the agreement between the experimental and the theoretical cross sections improves with the order of the ADC scheme. The double narrow peak at near 20 eV is reproduced by Stieltjes imaging as a single peak. The main difference between the ADC(2) and ADC(2)x results is in the region of the cross-section minimum that is described more accurately by the ADC(2)x scheme.

Finally, the oscillations present in the experimental cross section in the 30–60 eV range are missed completely by the ADC(1) result and they are not yet fully reproduced by the second order ADC(2) and ADC(2)x SI results. More specifically we observe that the first plateau around 35 eV is well reproduced by both second order methods, while the second one at 55 eV is not. This is due to the low resolution of the converged Stieltjes orders, interpolated to obtain the cross-sections, in the 45–65 eV energy range. Average relative deviations of the computed carbon dioxide cross-sections from the experimental one are given in Table II.

V. CONCLUSIONS

Stieltjes imaging has been long established as an efficient way of calculation of total photoionization cross-sections
using discretized continuum pseudospectra of the final states. However, the accuracy of this technique is limited by both the ability of the chosen \( L^2 \) basis set to represent continuum functions within the interaction volume and the numerical instability of the computational algorithm of the Stieltjes-Chebyshev moment theory. In view of these limitations, it could be doubted that improving the many-body theoretical description of the ionized system leads to significantly better cross-sections justifying the required higher numerical effort. Indeed, the resulting difference in the calculated cross-sections might fall within the margins of the inaccuracy incurred by the basis set and the Stieltjes imaging procedure. Our first results on ADC-Lanczos-Stieltjes method\(^{11} \) indicate that this is actually not the case and full inclusion of double electronic excitations does lead to more accurate Stieltjes imaging cross-sections. However, our initial work dealt only with two atomic and one molecular system. In the present paper, we have shown beyond doubt that the trend seen in Ref. 11 is characteristic of molecules of first row atoms in the valence energy region. Within the specific family of post-HF many-electron methods used here (ADC), ADC(2) leads to clear, substantial improvement over the single-excitation ADC(1) theory for all molecules considered, while for some of them, even a more demanding ADC(2)x level of theory leads to better agreement with the available experimental data (see Table II). On average, the precision gain achieved with ADC(2)x relative to ADC(2) in the considered energy window (from ionization threshold up to 100 eV) is about three times smaller than the precision gain of ADC(2) relative to the single-excitation ADC(1) method. It is instructive to analyze the relative deviations of the three \textit{ab initio} methods as function of photon energy (Fig. 9). Indeed, one observes that below 60 eV both ADC(2) and ADC(2)x methods lead to impressive agreement with experiment with the relative deviations below 10\%. At higher photon energies accuracy of all the ADC schemes grows reaching 20\% level around 80 eV. Since this behavior does not depend on the level of \textit{ab initio} theory, we conclude that it has to do with the limitations of the Gaussian single-electron basis sets.

The present work establishes the ADC-Lanczos-Stieltjes method as an efficient and reasonably accurate technique for molecular cross-sections in the valence region. Indeed, even within an unoptimized straightforward implementation of the method on the Intel Core i7-2600 processor, typical CPU time required for the cross-section calculations presented here is of the order of a few hours. Generalizations of the ADC-Lanczos-Stieltjes method to core ionization and to photoionization of excited molecular states are currently under development. Another promising direction for future research has to do with the most serious drawback of the Stieltjes-Chebyshev moment theory method, i.e., with its limited energy resolution. This can only be overcome by employing single-electron bases better adapted to bound-continuum transition calculations than the standard Gaussian ones. Specifically the B-spline based basis sets\(^{40} \) can provide the needed energy resolution and thus development of the B-spline techniques going beyond single excitations is most desirable. This, we believe, will also allow accurate description of the high-energy tails of the valence cross-sections.

**ACKNOWLEDGMENTS**

The authors would like to thank V. Vysotskiy for providing the interface for the two-electron integrals. V.A. and M.R. acknowledge the EU support through the Marie Curie ITN CORINF. V.A. acknowledges the financial support of the Engineering and Physical Sciences Research Council (EPSRC, UK) through the Career Acceleration Fellowship (Award No. EP/H003657/1) and the Programme Grant on Attosecond Dynamics (Award No. EP/I032517). V.A., K.G., and L.S.C. acknowledge the support of the Research Unit 1789 of the Deutsche Forschungsgemeinschaft.

---


